

ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN PELANGGAN MENGGUNAKAN PEMODELAN *LATENT DIRICHLET ALLOCATION* DAN *BERTopic* (STUDI KASUS: RESTORAN PAYAKUMBUAH GUBENG)

Azzam Pahlawan Ramadhan¹, Edbert Fernando²

¹2043221006, ²2043221106

^{1,2,3}Statistika Bisnis, Fakultas Vokasi, Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

2043221106@student.its.ac.id

Abstrak

Industri makanan dan minuman sebagai salah satu penyumbang PDB skala nasional mengalami pertumbuhan yang cukup signifikan. Kekuatan utama dari sektor ini adalah kearifan lokal dalam makanan tradisional, seperti nasi padang yang telah tersebar di seluruh Indonesia. Saat ini, hampir setiap wilayah di Indonesia terdapat usaha masakan padang. Penelitian ini akan menganalisis ulasan salah satu usaha masakan padang terbesar di Surabaya, yaitu Restoran Payakumbuh Gubeng. Penelitian ini dilakukan dengan melakukan pengambilan 1000 data ulasan terbaru dari *Google Rview*. Selanjutnya, dilakukan *pre-processing* data agar data dapat dianalisis lebih lanjut. Penelitian ini membandingkan dua metode, yaitu *Latent Dirichlet Allocation* dan *BERTopic*. Hasil analisis menunjukkan bahwa penentuan jumlah topik untuk model terbaik pada metode *BERTopic* yaitu sebanyak 10 topik, sedangkan pada metode LDA sebanyak 4 topik. Metode *BERTopic* memiliki model yang lebih baik dibandingkan metode LDA secara kuantitatif melalui nilai koherensi. Namun, metode *BERTopic* menghasilkan topik dengan kata kunci yang sering tumpang tindih dan sulit dipahami, sementara *Latent Dirichlet Allocation* (LDA) menghasilkan topik yang lebih jelas dan mudah dimengerti. Hasil pemodelan topik menggunakan LDA mencakup makanan, rasa, layanan, dan rekomendasi di Restoran Payakumbuh.

Kata kunci: Analisis Sentimen, *BERTopic*, LDA, Usaha Masakan Padang, Restoran Payakumbuh.

Article History

Received: July 2025
Reviewed: July 2025
Published: July 2025

Plagiarism Checker No 234
Prefix DOI :
10.8734/Musytari.v1i2.365
Copyright : Author
Publish by : Musytari



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)

I. PENDAHULUAN

Industri makanan dan minuman (F&B) merupakan salah satu sektor yang memegang peranan krusial dalam pertumbuhan ekonomi Indonesia. Menurut data Kementerian Perindustrian (Kemenperin, 2024), industri ini memberikan kontribusi sebesar 39,10 persen terhadap pendapatan domestik bruto (PDB) industri nonmigas serta 6,55 persen terhadap PDB nasional. Salah satu kekuatan utama industri F&B terletak pada kekayaan dan keragaman kuliner nusantaranya, dengan masakan Padang sebagai salah satu ikon yang paling menonjol. Skala industri ini sangat masif, terbukti dari estimasi 20.000 usaha rumah makan Padang yang beroperasi di area Jakarta dan sekitarnya saja, belum termasuk penyebarannya di kota-kota besar lain seperti Surabaya [1].

Besarnya skala pasar usaha masakan Padang ini secara langsung menciptakan lanskap persaingan yang cukup ketat. Hal tersebut menyebabkan usaha masakan Padang tidak hanya bersaing dengan usaha makanan lain, tapi juga bersaing antar *brand* masakan Padang. Dalam konteks ini, Restoran Payakumbuh di Surabaya sebagai studi kasus yang relevan karena memiliki jejak digital yang signifikan yang ditandai oleh volume ulasan yang cukup ramai di platform *Google Maps*. Volume ulasan yang besar ini mengindikasikan basis pelanggan yang luas sekaligus menunjukkan adanya beragam opini publik. Analisis awal terhadap ulasan-ulasan ini menunjukkan adanya sentimen yang bervariasi, mulai dari kepuasan hingga umpan balik kritis yang dapat dijadikan sebagai salah satu acuan bahan evaluasi bagi pihak manajemen restoran. Salah satu teknik dalam analisis sentimen yang dapat menganalisis persoalan tersebut adalah *topic modelling*.

Topic modelling atau pemodelan topik adalah teknik pembelajaran mesin (*machine learning*) tanpa pengawasan (*unsupervised*) untuk mengeksplorasi dan mengungkapkan struktur semantik dari sekumpulan dokumen berbentuk teks [2]. Terdapat beberapa metode yang umum digunakan dalam analisis *topic modelling*, di antaranya *Latent Semantic Analysis* (LSA), *Probabilistic Latent Semantic Analysis* (PLSA), *Latent Dirichlet Allocation* (LDA), *Correlated Topic Model* (CTM), dan *BERTopic* [3]. Namun, dalam penelitian ini akan dilakukan analisis menggunakan perbandingan metode LDA dan *BERTopic*. LDA merupakan metode untuk pemodelan topik adalah yang didasarkan kepada konsep probabilitas untuk mencari kemiripan suatu dokumen dan mengelompokkan dokumen-dokumen menjadi beberapa topik atau kelompok [4]. LDA dipilih karena perannya sebagai model dasar (*benchmark*) yang paling fundamental dan paling banyak dijadikan acuan dalam literatur *topic modeling*. Sedangkan *BERTopic* adalah model *state-of-the-art* yang mewakili generasi baru *topic modeling* [4].

Penelitian ini menggunakan data *review* terhadap Restoran Payakumbuh, Gubeng, Surabaya yang terdiri dari 1000 ulasan terbaru yang mencakup tingkat kepuasan dan kritik saran dari pelanggan. Analisis ini dimulai dengan *pre-processing* data dengan menghilangkan atau mengubah kata-kata yang tidak bernilai melalui proses *cleaning data*, *case folding*, *tokenizing*, *filtering*, dan *stemming*. Selanjutnya, data akan dibentuk menjadi *document term matrix* yang merupakan representasi matematis dari frekuensi kata-kata yang muncul dalam ulasan pelanggan. Matriks ini akan digunakan untuk mengevaluasi kualitas model pemodelan topik dengan menghitung nilai *coherence score*. Nilai *coherence score* tersebut akan digunakan untuk acuan awal sebelum dilakukan pemodelan. Selanjutnya, akan dilakukan pemodelan ulasan menggunakan metode LDA dan *BERTopic*.

II. TINJAUAN PUSTAKA

A. *Pre-Processing* Data

Pre-processing data merupakan proses mengolah data yang tidak terstruktur menjadi lebih struktur. Proses *pre-processing* data terdiri dari penyetaraan teks, menghapus angka-angka, tanda baca, karakter khusus, dan mengekstrak *stopwords*. Pertama, proses dilakukan dengan menghilangkan atau mengubah kata-kata yang tidak bernilai. Dalam proses ini semua kata akan diubah menjadi huruf kecil. Setelah itu, proses dilanjutkan dengan cara menghapus spasi, tanda baca, karakter khusus, dan *stopwords* karena umumnya hal tersebut tidak berkontribusi pada identifikasi topik [2].

B. *Document Term Matrix* (DTM)

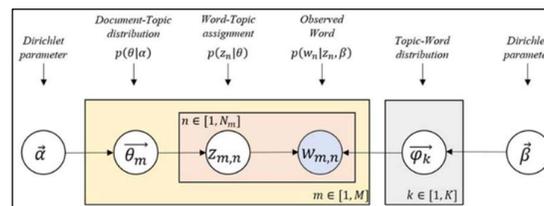
Document Term Matrix (DTM) merupakan suatu matriks yang mencerminkan hubungan antara dokumen dengan kata atau istilah (*term*). Setiap baris dalam DTM mewakili dokumen dan setiap kolom dalam DTM mewakili *term*, sementara nilai entrinya adalah jumlah frekuensi kemunculan *term* dalam dokumen [5]. Penggunaan DTM terbukti efektif dalam menyajikan pola distribusi kata yang mencerminkan struktur tematik tersembunyi di balik data teks tidak terstruktur [6].

C. Topic Modelling

Topic Modelling merupakan metode analisis teks yang bertujuan untuk mengelompokkan dokumen berdasarkan topik-topik tertentu. Teknik ini masuk dalam kategori *clustering* karena mengelompokkan dokumen berdasarkan tingkat kemiripan isi. Meskipun memiliki kesamaan tujuan dengan klasifikasi, pendekatan yang digunakan dalam *topic modelling* cukup berbeda. Fokus utama dari *topic modelling* yaitu untuk mengidentifikasi topik-topik abstrak yang tersembunyi dalam kumpulan dokumen, memberikan gambaran umum terhadap isi dokumen berdasarkan topik yang ditemukan, serta mengelompokkan dokumen berdasarkan tinjauan tersebut. Selain itu, metode ini juga bertujuan untuk menemukan Kumpulan kata kunci yang menjadi representasi dari masing-masing topik yang tersembunyi [6]. Hasil dari penerapan *topic modeling* berupa sejumlah topik yang terbentuk dari kumpulan kata yang sering muncul bersama dalam dokumen, berdasarkan pola kemunculan kata-kata tersebut [7].

D. Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) merupakan model generatif berbasis probabilistik yang digunakan untuk menganalisis data diskrit, seperti kumpulan teks. LDA termasuk dalam model Bayesian hierarkis tiga tingkat, dimana setiap elemen dalam kumpulan data dianggap sebagai hasil dari kombinasi beberapa topik. Tujuan utama dari metode LDA yaitu untuk mengidentifikasi jumlah topik dalam suatu korpus serta distribusi kata dalam masing-masing topik tersebut [8]. Untuk memperjelas cara kerja algoritma ini, ilustrasi pemodelan topik menggunakan LDA dapat dilihat pada Gambar 1.



Gambar 1. Pemodelan LDA (Sumber: Lee et all., 2018)

E. BERTopic

BERTopic merupakan metode pemodelan topik berbasis *embedding* yang memanfaatkan representasi semantik dari teks menggunakan model transformer yang bertujuan untuk menghasilkan topik yang lebih kontekstual dan mudah diinterpretasi. Tidak seperti LDA yang mengandalkan frekuensi kata dalam DTM, *BERTopic* mengelompokkan dokumen dengan mempertimbangkan makna kalimat secara keseluruhan, sehingga mampu menangkap topik-topik halus dalam data ulasan yang bersifat informal dan penuh variasi ekspresi, seperti yang umum ditemukan dalam ulasan pelanggan restoran di platform digital. Metode ini menggabungkan teknik *clustering* berbasis HDBSCAN dan reduksi dimensi seperti UMAP untuk menemukan struktur topik yang alami. Dalam banyak kasus, *BERTopic* memberikan hasil yang lebih komprehensif dibandingkan metode klasik, khususnya dalam konteks analisis sentimen dan analisis ulasan konsumen [9].

F. Coherence Score

Nilai koherensi (*coherence score*) merupakan ukuran yang digunakan untuk menilai seberapa konsisten atau mudah dipahami sekumpulan kata yang dihasilkan dari proses pemodelan topik. Melalui skor ini, peneliti dapat menentukan jumlah topik yang optimal berdasarkan kualitas interpretasi terhadap kata-kata dalam setiap topik. Biasanya, nilai koherensi divisualisasikan dalam bentuk grafik untuk memudahkan analisis. Selanjutnya, jumlah topik terbaik dipilih berdasarkan nilai koherensi tertinggi atau selisih antar skor yang signifikan. Semakin tinggi nilai koherensinya, maka semakin baik pula tingkat keterbacaan dan keterhubungan topik oleh manusia [10].

G. Usaha Masakan Padang

Masakan Padang merupakan salah satu representasi kuliner Nusantara yang memiliki popularitas tinggi di seluruh wilayah Indonesia. Ciri khas rasanya yang kuat, penggunaan rempah yang kompleks, serta sistem penyajian unik menjadikan usaha rumah makan Padang tidak hanya bertahan, tetapi juga terus berkembang. Keberhasilan rumah makan Padang terletak pada kemampuannya dalam menggabungkan resep warisan budaya dengan model bisnis yang adaptif terhadap preferensi konsumen lintas daerah [11]. Saat ini, usaha masakan Padang menghadapi lanskap persaingan yang ketat, terutama di kota-kota besar seperti Surabaya, Jakarta, dan Medan. Brand “Restoran Payakumbuh” menjadi salah satu usaha masakan padang yang besar dan telah tersebar di seluruh Indonesia.

III. METODOLOGI PENELITIAN

A. Sumber Data

Sumber data yang digunakan penelitian ini adalah data sekunder yang diperoleh dari 1000 ulasan terbaru Restoran Payakumbuh, Gubeng, Surabaya melalui *Google Riview*. Pengambilan data dilakukan pada hari Sabtu, 20 Juni 2025 di Kota Surabaya.

B. Langkah Analisis

Langkah analisis yang dilakukan dalam penelitian ini adalah sebagai berikut.

1. Melakukan pengambilan data 1000 ulasan terbaru Restoran Payakumbuh melalui *Google Maps*.
2. Melakukan *pre-processing* data yang terdiri dari *case folding*, *tokenizing*, penghapusan *stopwords*, penghapusan angka, penghapusan simbol, dan melakukan *stemming*.
3. Membuat *document term matrix* (DTM).
4. Menentukan jumlah topik optimum yang akan digunakan dalam pemodelan menggunakan *coherence score*.
5. Melakukan *topic modelling* menggunakan metode LDA dan *BERTopic*.
6. Melakukan perbandingan metode LDA dan *BERTopic*.
7. Menginterpretasikan hasil analisis.
8. Menarik kesimpulan dan saran.

IV. HASIL DAN PEMBAHASAN

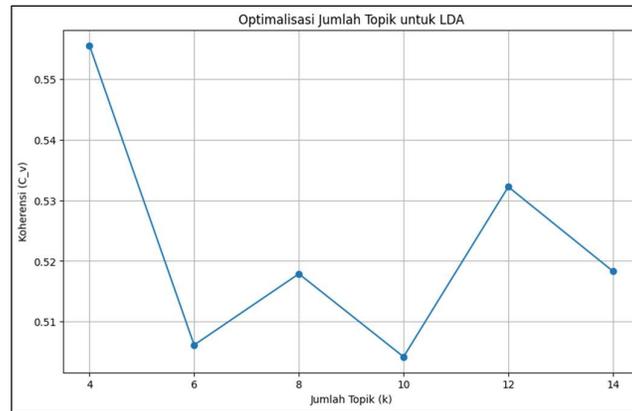
A. Pre-Processing Data

Pre-processing pada ulasan Restoran Payakumbuh Gubeng terdiri dari *case folding*, penghapusan *stopwords*, *stemming*, dan *tokenizing* yang selanjutnya akan dibentuk *document term matrix*. Hasil *pre-processing* data adalah sebagaimana ditunjukkan pada Tabel 1.

Tabel 1. *Document Term Matrix Pre-Processing Data*

| No | Enak | Makan | Mantap | Nasi | | Menu |
|------|------|-------|--------|------|------|------|
| 1017 | 2 | 1 | 1 | 0 | | 1 |
| 911 | 1 | 1 | 0 | 0 | | 1 |
| 798 | 2 | 0 | 3 | 1 | | 0 |
| : | : | : | : | : | : | : |
| 41 | 3 | 2 | 1 | 1 | : | 1 |

Tabel 1 menunjukkan struktur data setelah dilakukan *pre-processing* data. Data sudah berbentuk *document term matrix*, sehingga dapat dilakukan analisis selanjutnya.



Gambar 4. Plot Coherence Score

Gambar 4 menunjukkan bahwa *coherence score* pada setiap topik. Tahap awal pemodelan topik adalah dengan melakukan pemodelan menggunakan 14 topik. Hasil pemodelan tersebut menunjukkan bahwa banyak topik yang memiliki nilai *coherence score* paling besar adalah 10 topik dengan nilai *coherence score* sebesar 0,56. Dengan begitu, pemodelan yang dilakukan yaitu dengan menggunakan 4 topik.

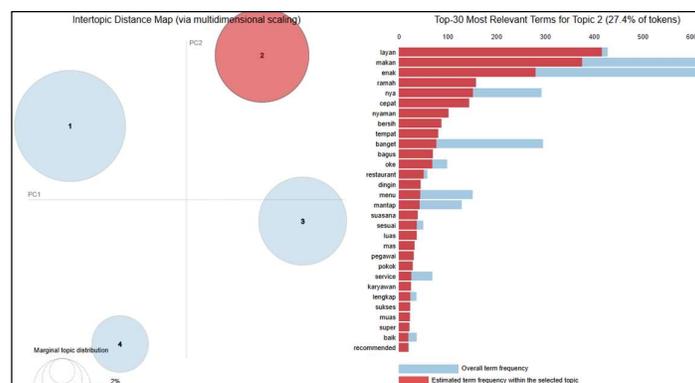
D. Latent Dirichlet Allocation (LDA)

Hasil pemodelan topik menggunakan *Latent Dirichlet Allocation* (LDA) menggunakan LDA dengan topik berjumlah 10 adalah sebagaimana ditunjukkan pada Tabel 2.

Tabel 2. Hasil Pemodelan Topik LDA

| Topik | Terms | Interpretasi |
|-------|--|--------------|
| 1 | Nasi, ayam, telur, pesan, tunggu, makan, antri, saji, rendang, barendo, meja, orang, bayar | Makanan |
| 2 | Layan, makan, enak, ramah, cepat, nyaman, bersih, tempat, bagus | Pelayanan |
| 3 | Enak, padang, makan, coba, Surabaya, Payakumbuh, bumbu, nasi | Rasa |
| 4 | Good, dendeng best, service, food, padang, recommended | Rekomendasi |

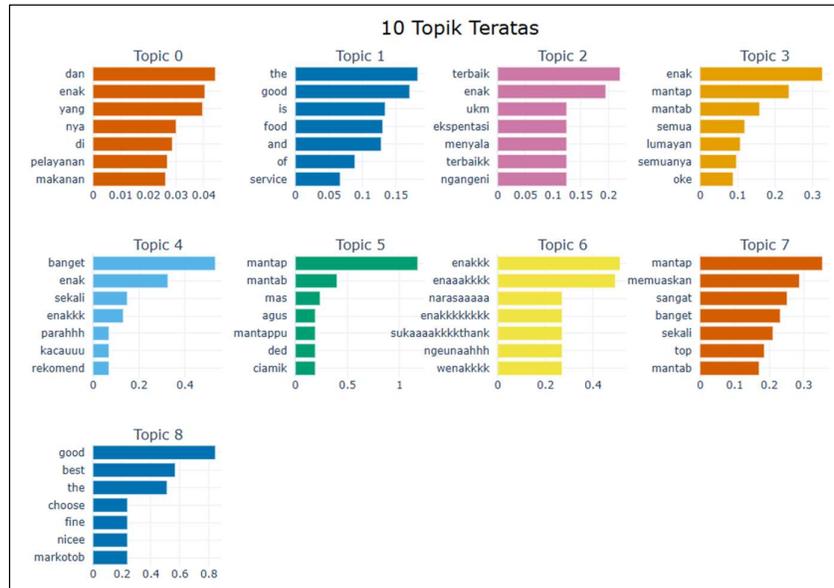
Tabel 2 menunjukkan hasil pemodelan topik menggunakan metode LDA. Topik 1 memiliki kata-kata kunci yang merujuk pada makanan yang terdapat di Restoran Payakumbuh. Topik 2 mencerminkan pelayanan yang diberikan oleh restoran. Topik 3 merujuk pada rasa makanan pada restoran Payakumbuh. Topik 4 menunjukkan kata-kata kunci yang merujuk pada rekomendasi agar khalayak mempertimbangkan memilih restoran ini sebagai tempat makan. Visualisasi hasil pemodelan topik menggunakan model LDA adalah sebagaimana ditunjukkan pada Gambar 5.



Gambar 5. Visualisasi Model LDA

E. BERTopic

Hasil pemodelan topik menggunakan metode *BERTopic* dilakukan dengan jumlah 10 topik adalah sebagaimana Gambar 6.



Gambar 6. Visualisasi Pemodelan Topik *BERTopic*

Gambar 6 menunjukkan bahwa hasil visualisasi dari pemodelan 10 topik menggunakan *BERTopic*. Kata yang paling sering muncul pada topik-topik tersebut yaitu “enak”, “mantap”, “good”, “best”, dan “terbaik”.

F. Perbandingan Metode LDA dan *BERTopic*

Perbandingan pemodelan topik menggunakan LDA dan *BERTopic* dapat dilihat melalui perbandingan nilai koherensi yang ditunjukkan sebagaimana Tabel 3.

Tabel 3. Perbandingan Nilai Koherensi

| Metode | Koherensi |
|-----------------|-----------|
| <i>BERTopic</i> | 0,5441 |
| LDA | 0,5226 |

Tabel 3 menunjukkan bahwa nilai koherensi metode *BERTopic* sebesar 0,5541, serta nilai koherensi metode LDA yaitu sebesar 0,5226. Hal tersebut menunjukkan bahwa metode *BERTopic* memiliki nilai koherensi yang lebih besar dari metode LDA, sehingga dapat disimpulkan bahwa *BERTopic* lebih baik daripada LDA secara kuantitatif. Namun, pada penerapannya, metode LDA cenderung memiliki kata kunci yang lebih jelas dan terfokus, sehingga memudahkan dalam penafsiran dan pemahaman topik yang dihasilkan. Maka dari itu, meskipun *BERTopic* lebih unggul secara metrik kuantitatif, namun untuk tujuan identifikasi aspek sosial dan bisnis, model LDA dalam kasus ini cenderung lebih efektif.

V. KESIMPULAN DAN SARAN

Penentuan jumlah topik optimum menggunakan *coherence score* menunjukkan bahwa model dengan 10 topik pada metode *BERTopic* dan 4 topik pada metode LDA memiliki kualitas terbaik. Berdasarkan nilai *coherence score* juga didapatkan bahwa model *BERTopic* lebih baik. Namun, metode *BERTopic* menghasilkan topik dengan kata kunci yang sering tumpang tindih dan sulit dipahami, sementara *Latent Dirichlet Allocation* (LDA) menghasilkan topik yang lebih

jelas dan mudah dimengerti. Hasil pemodelan topik menggunakan LDA mencakup makanan, rasa, layanan, dan rekomendasi di Restoran Payakumbuh . Saran untuk penelitian selanjutnya yaitu sebaiknya lebih mengeksplorasi penggunaan model lain seperti *Latent Semantic Analysis* (LSA), *Probabilistic Latent Semantic Analysis* (PLSA). Selain itu, penulis selanjutnya juga disarankan untuk melakukan perbandingan metode yang setara agar lebih akurat dalam mendapatkan hasil analisis.

DAFTAR PUSTAKA

- [1] D. Darmawan and M. Firmansyah, *Bisnis Antbangkrut: Membongkar Rahasia Sukses Rumah Makan Padang & Warteg*. Jakarta: Transmedia Pustaka, 2019.
- [2] Alghamdi, R., & Alfalqi, K. "A Survey of Topic Modeling in Text Mining". *International Journal of Advanced Computer Science and Applications*.
- [3] T. Kwartler, "Text mining in practice with R", vol. 90, no. 13 (2017)
- [4] J. Yoon-Hwang, W.-R. Yoo, and Y.-J. Park, "Research Trend Analysis Using BERTopic and LDA," *The Journal of Digital Contents Society*, vol. 24, no. 5, pp. 857-864, May 2023.
- [5] Pradana, M. G. "Penggunaan fitur wordcloud dan document term matrix dalam text mining". *Jurnal Ilmiah Informatika*, 8(01), 38-43 (2020).
- [6] N. Ahmed, M. A. Khuhro, and M. A. Dootio, "Building and analyzing a Brahui text corpus: Applying DTM and TF-IDF techniques," *ResearchGate*, 2024.
- [7] C. Naury, D. H. Fudholi, and A. F. Hidayatullah, "Topic Modelling pada Sentimen Terhadap Headline Berita Online Berbahasa Indonesia Menggunakan LDA dan LSTM," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 24, 2021
- [8] F. E. Priady and I. Irhamah, "Sistem Rekomendasi Buku Bacaan untuk Anak Menggunakan Collaborative Filtering dan Topic Modelling," *Jurnal Sains dan Seni ITS*, vol. 14, no. 1, pp. 1-8, 2025.
- [9] J. Yoon-Hwang, W.-R. Yoo, and Y.-J. Park, "Research Trend Analysis Using BERTopic and LDA," *J. Digit. Contents Soc.*, vol. 24, no. 5, pp. 857-864, May 2023.
- [10] D. F. Simanjuntak, "Perbandingan Kinerja Word Embedding dan Pemodelan Topik Dalam Identifikasi Topik Tugas Akhir Berdasarkan Judul dan Abstrak Penelitian", Skripsi S1, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia, 2025.
- [11] D. Darmawan and M. Firmansyah, *Bisnis Antbangkrut: Membongkar Rahasia Sukses Rumah Makan Padang & Warteg*. Jakarta: Transmedia Pustaka, 2009.